IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

TAKE2 TECHNOLOGIES LIMITED, a Cayman Islands company; and THE CHINESE UNIVERSITY OF HONG KONG, a Hong Kong entity, Plaintiffs, v. PACIFIC BIOSCIENCES OF CALIFORNIA, INC., a Delaware corporation Defendant.	C.A. No.: **JURY TRIAL DEMANDED** **PUBLIC VERSION FILED ON: DECEMBER 14, 2022**

COMPLAINT

1. Plaintiff Take2 Technologies Limited ("Take2") and Nominal Plaintiff The Chinese University of Hong Kong ("CUHK") (collectively, "Plaintiffs") file this Complaint against Defendant Pacific Biosciences of California, Inc. ("PacBio"), alleging as follows:

NATURE OF THE ACTION

2. This is an action for infringement of U.S. Patent No. 11,091,794 ("the '794 Patent" or "the Take2 Patent") arising under the patent laws of the United States, 35 U.S.C. §§ 100 *et seq.*, including 35 U.S.C. § 271. A true and correct copy of the '794 Patent is attached hereto as Exhibit A. PacBio has infringed and continues to infringe the claims of the '794 Patent by making, using, offering for sale, and selling its DNA sequencing platforms, systems, software, and services in the United States.

PARTIES

3. Plaintiff Take2 is a company organized and existing under the laws of the Cayman Islands, with its place of business at Portcullis (Cayman) Ltd, Grand Pavilion Commercial Centre, Oleander Way, 802 West Bay Road, P.O. Box 32052 Grand Cayman, KY1-1208 Cayman Islands.

4. Nominal Plaintiff CUHK is a university established by legislation in the Hong Kong Special Administrative Region ("Hong Kong SAR"), located at Shatin, the New Territories, Hong Kong SAR.

5. Plaintiffs are informed and believe that: Defendant Pacific Biosciences of California, Inc. is a corporation organized and existing under the laws of the State of Delaware, has its principal place of business at 1305 O'Brien Drive, Menlo Park, California, 94025, USA.

JURISDICTION AND VENUE

6. This is a civil action for patent infringement arising under the patent laws of the United States, 35 U.S.C. §§ 100, *et seq.*, and in particular, 35 U.S.C. § 271.

7. This Court has subject matter jurisdiction over this patent infringement action under 28 U.S.C. §§ 1331 and 1338(a).

8. This Court has personal jurisdiction over PacBio because PacBio is incorporated in Delaware and has purposely availed itself of the benefits and protections of Delaware's laws such that it should reasonably anticipate being sued in this Court.

9. Upon information and belief, PacBio, itself and through its affiliates and service providers, develops, manufactures, imports, markets, distributes, tests, uses, offers to sell, and/or sells its products and services related to DNA sequencing throughout the United States, including in the District of Delaware, and therefore transacts business within the District of Delaware, and/or has engaged in systematic and continuous business contacts within the District of Delaware.

10. Venue lies in this judicial district pursuant to 28 U.S.C. §§ 1391(b) and (c), and 1400(b).

BACKGROUND

The ’794 Patent

11. The technology at issue in this case relates to the study of “epigenetics.” The basic structure of our genes is formed from a backbone of DNA molecules, the sequence of which is made up of four different structural units or bases called nucleotides: cytosine (C), adenine (A), thymine (T) and guanine (G). It is generally well known that the sequence of one’s genes are inherited. Genetics, therefore, is the study of how certain traits such as disease or other health conditions are passed from parents to offspring as a result of changes in a gene’s DNA sequence. In contrast, the field of epigenetics studies changes to a DNA molecule caused by chemical modifications to the nucleotide subunits of DNA rather than changes to the base sequence itself. Epigenetic changes to DNA have become an important focus in studying why certain individuals with the same genetic background can have very different susceptibilities to disease, because such modifications can be initiated or disrupted by environmental factors such as stress, diet, aging, drug use, and pollutants. Epigenetic changes also contribute to the differences in gene expression between cells within a single individual when the DNA sequences of these cells are the same. Thus, understanding the epigenome holds important clues to one’s current and future health status and one day may allow for better prevention, diagnosis, and treatment of disease or other health conditions.

12. There are several types of modifications that can give rise to an epigenetic change, one of which is DNA methylation. Methylation involves the addition of a methyl group to one or more nucleotides, C, A, T, or G. Methylation most often occurs within regions of DNA called CpG sites, where a cytosine nucleotide is followed by a guanine nucleotide. Methylation of that cytosine forms 5-methylcytosine (5mC), which has been reported to be involved in the

development of cancer and other conditions.

13. Although there has been a need to develop techniques for measuring 5mC to study its role in the development and diagnosis of disease, existing techniques prior to filing the '794 Patent had failed to deliver sufficiently accurate results. The founders of Take2 and their team at CUHK—Yuk-Ming Dennis Lo, Rossa Wai Kwun Chiu, Kwan Chee Chan, Peiyong Jiang, Suk Hang Cheng, Wenlei Peng, and On Yee Tse (together, "the Take2/CUHK Team")—embarked on a plan to change that. To do so, the Take2/CUHK Team invented a new neural-network-based model they coined the "holistic kinetic" (HK) model. That HK model was developed using aligned sequence reads generated with a PacBio sequencing platform with information related to the context of each of a series of nucleotides along a string of nucleotides and how quickly each nucleotide is incorporated into a DNA strand to provide a highly accurate probability of methylation for each nucleotide of a sample window of DNA. The information for each nucleotide within such a sample window had not previously been collected and improves the ability of existing systems, like the PacBio sequencing platform, to detect nucleotide modifications such as 5mC methylation within a single set of sequence reads. The ability to detect nucleotide modifications is just one of the ground-breaking innovations by the Take2/CUHK Team led by Dr. Lo, who recently received the 2022 Lasker~DeBakey Clinical Medical Research Award for his discovery of fetal DNA in maternal blood and its use for noninvasive prenatal testing for Down syndrome. The invention for detecting nucleotide modifications is disclosed, and aspects of it claimed, in the '794 Patent.

The History Between PacBio and the Take2/CUHK Team

14. On January 25, 2021, the HK model was published in a research article in the journal Proceedings of the National Academy of Sciences (PNAS) entitled "Genome-wide

detection of cytosine methylation by single molecule real-time sequencing" ("the PNAS article"). Exhibit B (available at https://www.pnas.org/doi/10.1073/pnas.2019768118). Page 1 of the PNAS article contains a statement that notifies readers that a patent application was filed on the technology described in the article.

15. As customers of PacBio, the Take2/CUHK Team sent the PNAS article to PacBio representatives because the Take2/CUHK Team thought that PacBio would be excited to learn of the Take2/CUHK Team's discovery using the PacBio sequencing platform and hoped to generate interest from PacBio to negotiate a collaboration or licensing agreement surrounding that discovery. The response was indeed enthusiastic. Within a day of receiving the publication, several PacBio employees reached out to the Take2/CUHK Team via email to congratulate them on such a great piece of work and that it had generated significant interest within the company. PacBio even provides a link to access the PNAS paper on the PacBio Website at the location shown in annotated Fig. 1 below.

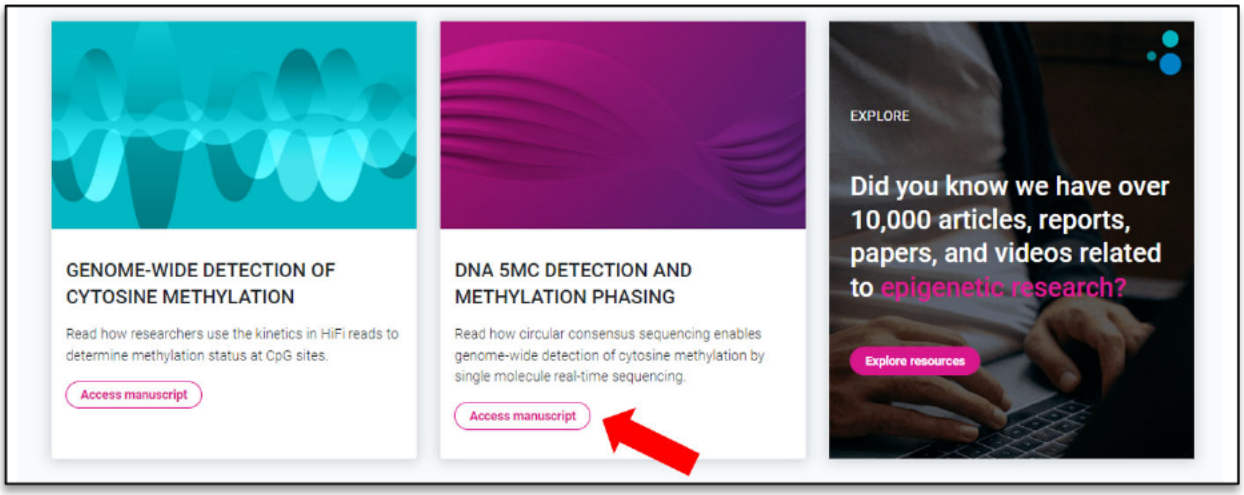


Fig. 1. Screenshot from https://www.pacb.com/products-and-services/applications/epigenetics/ (annotated) (Exhibit F)

16. On or about January 28, 2021, investors working on behalf of Take2 from Decheng Capital, Min Cui, Ph.D. and Victor E. Tong., Jr., [redacted]

[redacted]

17. [redacted], PacBio's President and CEO Christian Henry credited Dr. Dennis Lo for the "development of a new method for highly accurate simultaneous determination of DNA sequence and CPG [sic] methylation in one go, published in the PNAS journal in January. This new method will help researchers explore the impact of epigenetic changes in humans and other organisms, and has great potential in clinical research and eventually, diagnostics in cancer and other disease areas where methylation changes are known to be important disease markers" during its Q4 2020 Earnings call on February 10, 2021. Exhibit C at 4 (available at https://www.fool.com/earnings/call-transcripts/2021/02/10/pacific-biosciences-of-california-pacb-q4-2020-ear/).

18. [redacted]

19. [redacted]

[redacted]

20. [redacted]

21. [redacted]

22. [redacted]

23. On October 18, 2021, PacBio hosted a workshop at the Annual Meeting of the American Society of Human Genetics (ASHG). During the PacBio Workshop, PacBio's Vice

President of Segment Marketing, Jennifer Stone, highlighted "the great work coming out of Dennis Lo's lab at the Chinese University of Hong Kong earlier this year. The team developed a methodology to directly examine 5-methyl-C at the whole genome level using the kinetic information from real-time HiFi sequencing without the need for bisulfite or enzymatic conversions. In the PNAS paper, Dr. Lo and his team described the development of algorithms that leverage HiFi sequencing information from both DNA strands to improve the detection of 5-methyl-C. The paper describes its application to human whole-genome sequencing and the benefits of characterizing the epigenome with long accurate reads. It allows epigenetic phasing and allelic differential methylation. So just like the phasing of genetic variants, the ability to detect the methylation status genome-wide with long HiFi reads also allows for long-range epigenetic phasing information." PacBio, *HiFi Sequencing: See What You've Been Missing*. YOUTUBE (Nov. 2, 2021) at 05:05–06:06, https://www.youtube.com/watch?v=d2an7LXhPSU. (Also accessible at https://www.pacb.com/videos/hifi-sequencing-see-what-youve-been-missing/).

24. Upon information and belief, PacBio's Associate Director of Product Marketing, Aaron Wenger, Ph.D., also gave a presentation to the American Society of Human Genetics (ASHG) during its virtual annual meeting that was held on October 18-22, 2021 (the "ASHG presentation"). PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021), https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s. (Also accessible at https://www.pacb.com/videos/methylation-detection-with-pacbio-hifi-sequencing/).

25. During the ASHG presentation, Dr. Wenger discussed alleged improvements that PacBio planned to make to the PacBio sequencing platform, including "models that integrate kinetic information over multiple bases and passes of a molecule" that "can provide high precision

and recall for 5mC from PacBio HiFi reads." PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 15:46–15:56, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s. Dr. Wenger further explained that "there are tools available to call 5mC that are under development at PacBio and available to early access collaborators." PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 16:10–16:17, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s.

26. [redacted].

27. On or about April 20, 2022, PacBio released a new version of its SMRT Analysis software suite—SMRT Link v11.0—which has been designed for use with Single Molecule, Real-Time (SMRT) sequencing data generated by PacBio's Sequel II and Sequel IIe products. In the release notes, PacBio describes several new features of its software, including a "[n]ew data utility to perform **5mC CpG Detection**." Exhibit D at 2.

28. In a press release dated April 21, 2022, PacBio announced that Children's Mercy Kansas City was one of its first customers to use the methylation detection capability of its products running SMRT Link v11.0. *See* Exhibit E (available at https://www.pacb.com/press_releases/pacbio-and-childrens-mercy-kansas-city-expand-collaboration-taking-a-multi-omics-approach-to-characterize-rare-disease/).

29. [redacted].

30. [redacted]

CAUSE OF ACTION

(Infringement of U.S. Patent No. 11,091,794)

31. The Plaintiffs incorporate by reference the allegations contained in paragraphs 1–30 above.

32. Nominal Plaintiff CUHK is the owner of the entire right, title, and interest in and to the '794 Patent, which is entitled "Determination of Base Modifications of Nucleic Acids." Exhibit A. CUHK has granted Plaintiff Take2 an exclusive license to the '794 Patent, including the right to sublicense the '794 Patent. The '794 Patent was duly and legally issued by the United States Patent and Trademark Office on August 17, 2021, and is valid, subsisting, and in full force and effect.

33. The '794 Patent generally relates to methods of identifying nucleotide modifications such as methylation. Claim 1 of the patent recites:

> A method for detecting a modification of a nucleotide in a nucleic acid molecule, the method comprising:
> (a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in a sample nucleic acid molecule and
> obtaining, from the data, values for the following properties:
> for each nucleotide:
> an identity of the nucleotide,
> a position of the nucleotide within the sample nucleic acid molecule,
> a width of the pulse corresponding to the nucleotide, and

an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

(b) creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:

the identity of the nucleotide,

a position of the nucleotide with respect to a target position within the window,

the width of the pulse corresponding to the nucleotide, and

the interpulse duration;

(c) inputting the input data structure into a model, the model trained by:

receiving a first plurality of first data structures, each first data structure of the first plurality of data structures corresponding to a respective window of nucleotides sequenced in a respective nucleic acid molecule of a plurality of first nucleic acid molecules, wherein each of the first nucleic acid molecules is sequenced by measuring pulses in the optical signal corresponding to the nucleotides, wherein the modification has a known first state in a nucleotide at a target position in each window of each first nucleic acid molecule, each first data structure comprising values for the same properties as the input data structure,

storing a plurality of first training samples, each including one of the first plurality of first data structures and a first label indicating the first state of the nucleotide at the target position, and

optimizing, using the plurality of first training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels of the first labels when the first plurality of first data structures is input to the model, wherein an output of the model specifies whether the nucleotide at the target position in the respective window has the modification,

(d) determining, using the model, whether the modification is present in a nucleotide at the target position within the window in the input data structure.

34. PacBio has infringed, and continues to infringe, the ’794 Patent in violation of at least 35 U.S.C. § 271(a) and (b) by, itself and/or through its agents and end users (e.g., customers, distributors, and service providers), unlawfully and wrongfully making, using, testing,

manufacturing, importing, offering to sell, and/or selling DNA sequencing systems, analysis software and tools that perform one or more of the methods claimed in the '794 Patent, within, from and/or into the United States without permission or license from Plaintiffs, and will continue to do so unless enjoined by this Court.

35. Exemplary sequencing products, analysis software and tools that directly infringe the '794 Patent include, but are not limited to, Sequel® II systems, Sequel IIe Systems, and Revio™ Systems that are equipped with or otherwise used with SMRT® Link software v11.0, v11.1, or later, or are otherwise able to detect methylation (collectively, "the PacBio Products"). Upon information and belief, PacBio directly infringes the '794 Patent at least by making and manufacturing the PacBio Products, testing the PacBio Products, and performing quality control for the PacBio Products.

36. PacBio has infringed and continues to infringe at least claim 1 of the '794 Patent, literally or under the doctrine of equivalents. As discussed in paragraphs 37 to 63 below, the PacBio Products practice, either literally or under the doctrine of equivalents, each step recited in claim 1 of the '794 Patent.

37. *Preamble*. Upon information and belief, the PacBio Products practice the preamble of claim 1 of the '794 Patent, which recites "A method for detecting a modification of a nucleotide in a nucleic acid molecule..." For example, PacBio's website advertises that the PacBio Products can be used to perform "HiFi sequencing" to "explore DNA modifications" to "measure 5mC methylation" as shown in annotated Fig. 2 below.

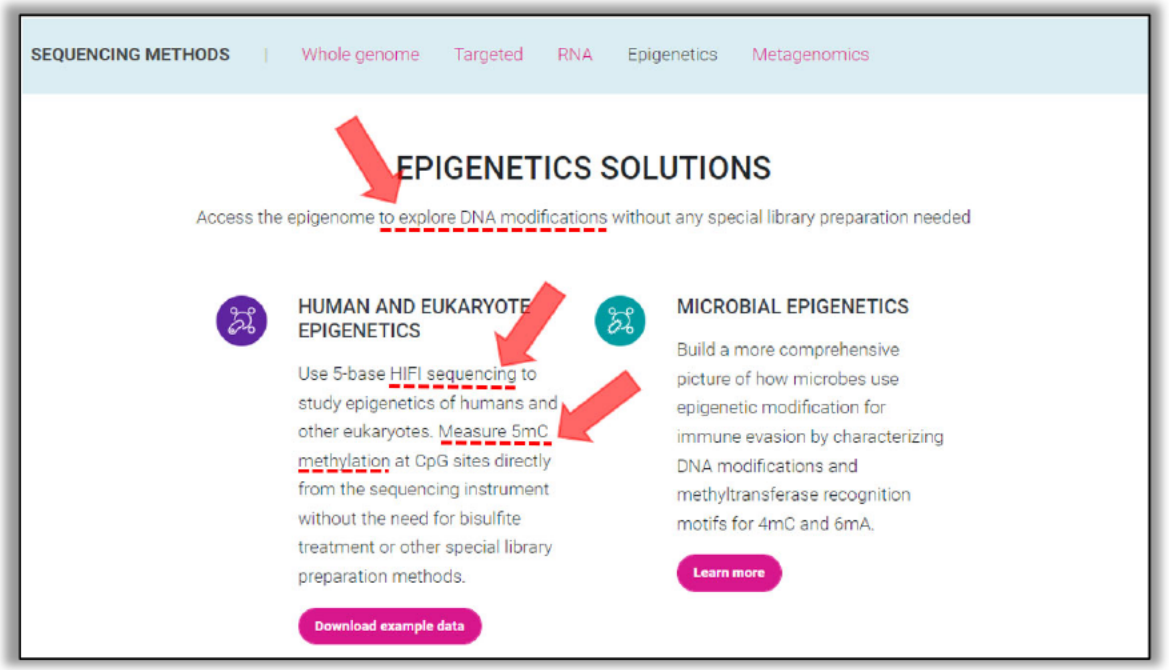


Fig. 2. Screenshot from https://www.pacb.com/products-and-services/applications/epigenetics/ (annotated) (Exhibit F)

38. PacBio's website also indicates that its "HiFi sequencing detects modifications in native DNA" to include "[m]ethylation detection with HiFi sequencing" as shown in annotated Fig. 3 below.

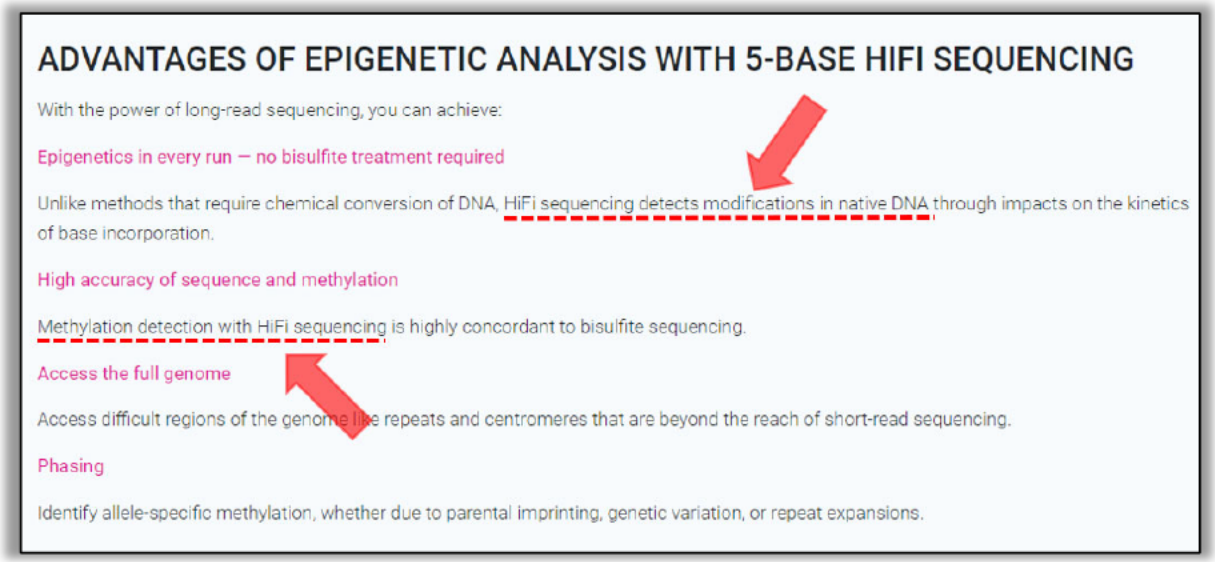


Fig. 3. Screenshot from https://www.pacb.com/products-and-services/applications/epigenetics/ (annotated) (Exhibit F)

39. Consequently, PacBio's HiFi sequencing method can be used to detect "a modification" (e.g., 5mC methylation) "in a nucleic acid molecule" (e.g., DNA) as claimed.

40. *Step (a)*. Upon information and belief, the PacBio Products practice step (a) of claim 1 of the '794 Patent, which recites:

> (a) receiving data acquired by measuring pulses in an optical signal corresponding to nucleotides sequenced in a sample nucleic acid molecule and obtaining, from the data, values for the following properties:
> for each nucleotide:
> an identity of the nucleotide,
> a position of the nucleotide within the sample nucleic acid molecule,
> a width of the pulse corresponding to the nucleotide, and
> an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;

41. PacBio's Sequel II and Sequel IIe systems (together, "the Sequel II systems") receive sequencing data generated from its HiFi sequencing technology. As shown in annotated Fig. 4 below, PacBio's website explains that "HiFi sequencing observes a polymerase incorporating fluorescently labeled nucleotides complementary to a native DNA strand" and that observation gives rise to the measurement of "two channels of information: fluorescence and kinetics."

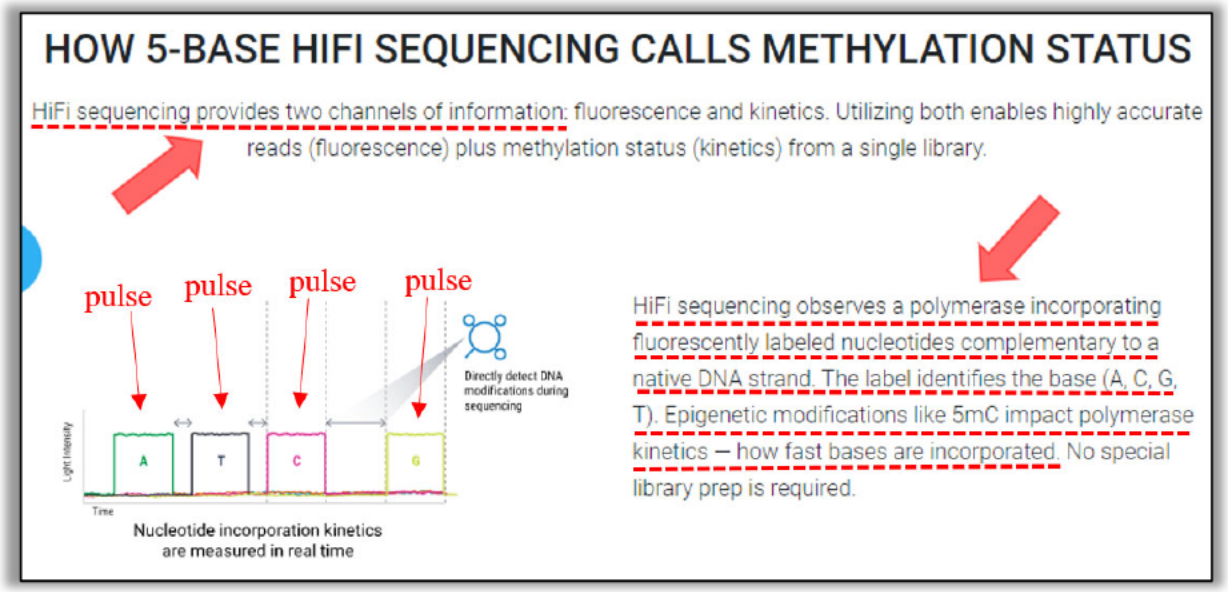


Fig. 4. Screenshot from https://www.pacb.com/products-and-services/applications/epigenetics/ (annotated) (Exhibit F)

42. The first measurement relates to fluorescently labeled nucleotides, wherein "[t]he label identifies the base (A, C, G, T)." Fig. 4 (Exhibit F). As Dr. Wenger explained in his presentation to ASHG, the fluorescent labels relate to "the **color** of the **pulses** that are observed" which "indicate[s] the identity of the base A, C, G, or T." PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 02:45–02:54, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s. Indeed, as shown in Fig. 4 above, each pulse is represented by a different fluorescent color to indicate a different base—green (A), black (T), pink (C), and yellow (G).

43. The second measurement relates to the kinetics, i.e., "how fast bases are incorporated" into the DNA strand. Polymerase kinetics are impacted by epigenetic modifications like 5mC. *See* Fig. 4 (Exhibit F). As explained by Dr. Wenger, "kinetics of the polymerase—how long it takes to incorporate a base and how long it goes between adjacent incorporations—is affected by both the context of the base and epigenetic modifications like methylation." PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at

02:54–03:10, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s. This is illustrated in Fig. 4 by the distance between pulses, for example, the time between the A (green) and T (black) is shorter than the time between C (pink) and G (yellow).

44. For the reasons discussed above, the sequencing data received by the Sequel II systems is "acquired by measuring pulses in an optical signal" (i.e., fluorescent pulses and kinetics) that "correspond[] to nucleotides sequenced in a sample nucleic acid molecule" (i.e., color of pulses and time between pulses that indicate the identity of and modifications to the nucleotide sequenced, respectively) as required by step (a) of claim 1.

45. Further, upon receiving the fluorescent pulses as discussed above, PacBio's Sequel II systems obtain the following values or information from the sequencing data for each nucleotide as required by claim 1:

Property required by claim 1	**Value**
"an identity of the nucleotide,"	The color of each fluorescent pulse corresponds to the identity of the nucleotide. (*see supra* ¶42, Fig. 4)
"a position of the nucleotide within the sample nucleic acid molecule,"	A subsequence or subread of a sample nucleic acid—referred to as a "query"— is aligned with a target or reference sequence, thereby obtaining the position of each nucleotide within the sample nucleic acid. *See* Exhibit G at 10 (PacBioFileFormats Documentation Release 11.0.0, available at https://pacbiofileformats.readthedocs.io/_/downloads/en/11.0/pdf/). That query provides

Property required by claim 1	Value
	a context for each nucleotide.
"a width of the pulse corresponding to the nucleotide, and"	The value of the width of the pulse corresponding to the nucleotide is indicated by the time it takes to incorporate a base. (*see supra* ¶43, Fig. 4)
"an interpulse duration representing a time between the pulse corresponding to the nucleotide and a pulse corresponding to a neighboring nucleotide;"	The value of the interpulse duration is indicated by the time between adjacent incorporations. (*see supra* ¶43, Fig. 4)

46. The values obtained above are then used to create an input data structure used in step (b) of claim 1, discussed below. Thus, the Sequel II systems practice every limitation of step (a) for the aforementioned reasons.

47. *Step (b)*. Upon information and belief, the PacBio Products also practice step (b) of claim 1 of the '794 Patent, which recites:

> (b) creating an input data structure, the input data structure comprising a window of the nucleotides sequenced in the sample nucleic acid molecule, wherein the input data structure includes, for each nucleotide within the window, the properties:
> the identity of the nucleotide,
> a position of the nucleotide with respect to a target position within the window,
> the width of the pulse corresponding to the nucleotide, and
> the interpulse duration;

48. The PacBio Products create an input data structure. For example, the PacBio Products first receive a file or set of files in BAM format that contain a plurality of HiFi reads obtained from the sequencing data ("the BAM file"). A BAM file contains sequence alignment data with specifications that are standardized so that the information provided in the fields are

generally applicable to raw or aligned sequence reads across different sequencing platforms. A working group maintains the industry standard fields, which is available at http://samtools.github.io/hts-specs/SAMv1.pdf. PacBio BAM files are compatible with the industry standard specifications. *See* Exhibit G at 5. The HiFi reads in the BAM file are then used to create the data structure(s) in step (b) that are used as an input—i.e., an input data structure—for step (c), as discussed below.

49. To create the HiFi reads in BAM format for the BAM file(s), the PacBio Products perform an analysis called circular consensus sequencing (CCS). *See, e.g.*, Exhibit H (available at https://www.pacb.com/technology/hifi-sequencing/) ("HiFi reads are produced using circular consensus sequencing (CCS) mode on PacBio long-read systems."); *see also* Exhibit I at 90–91, 104–106 (SMRT® Link user guide (v11.0), available at https://www.pacb.com/wp-content/uploads/SMRT_Link_User_Guide_v11.0.pdf); Exhibit J at 7–15 (SMRT® Tools reference guide (v11.0), available at https://www.pacb.com/wp-content/uploads/SMRT_Tools_Reference_Guide_v11.0.pdf); Exhibit L at 8–16 (SMRT® Tools reference guide (v11.1), available at https://www.pacb.com/wp-content/uploads/SMRT_Tools_Reference_Guide_v11.1.pdf); Exhibit M at 99–100, 115–117 (SMRT® Link user guide (v11.1), available at https://www.pacb.com/wp-content/uploads/SMRT_Link_User_Guide_v11.1.pdf).

50. The SMRT® Tools reference guide (v11.0) explains that the CCS process includes a "windowing" process in which "the subread-to draft alignment"—i.e., the nucleotides sequenced in the sample nucleic acid molecule—is divided into "overlapping *windows* with a target size of 22 bp with ±2 bp overlap." Exhibit J at 8 (emphasis added); *see also* Exhibit L at 9. Prior to the release of the SMRT® Tools reference guide (v11.0), PacBio used a 16-base-pair window of nucleotides for what it refers to as a "feature vector." For example, as explained by

Dr. Wenger during the development of this windowing process, every CpG site in a HiFi read is used to produce "a feature vector that consists of the kinetics, the pulse widths, and interpulse durations in a 16-base window around that CpG site on both strands of the sequence." PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 07:25–07:47, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s. Each feature vector, which is a smaller data structure built from the BAM file, serves as an input to a convolution neural network (CNN) model, which is discussed further below and is shown in Fig. 5 below:

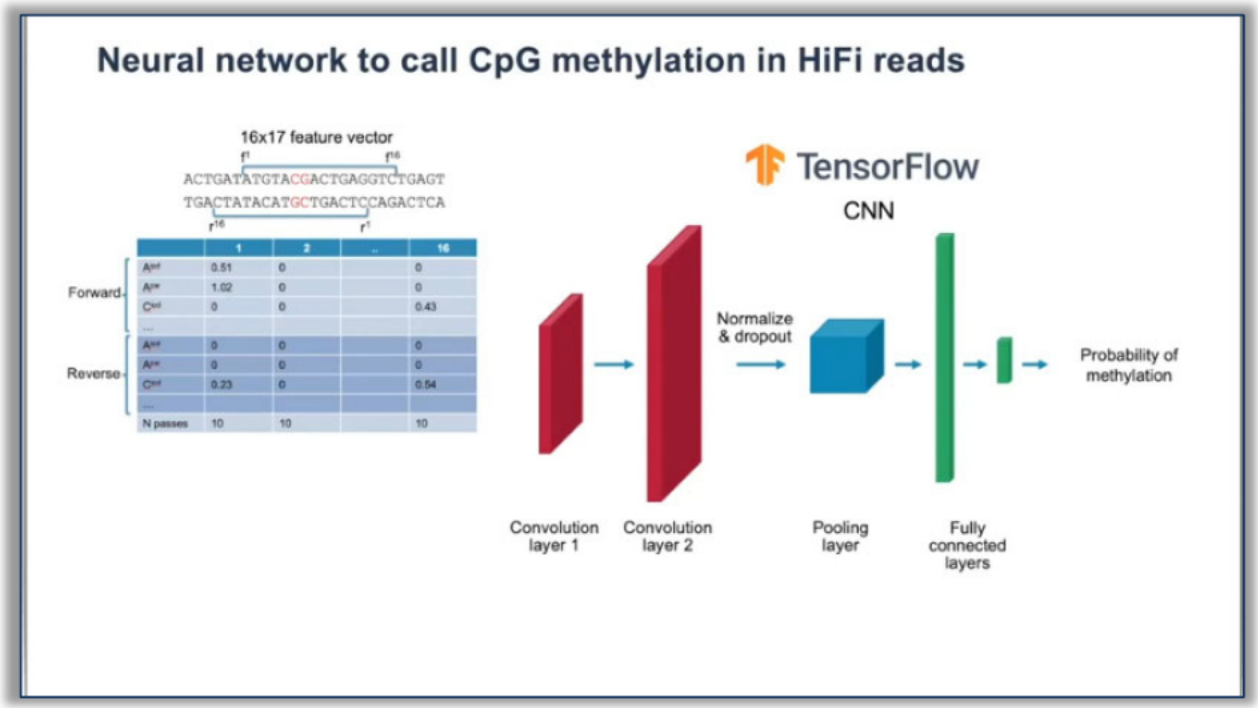
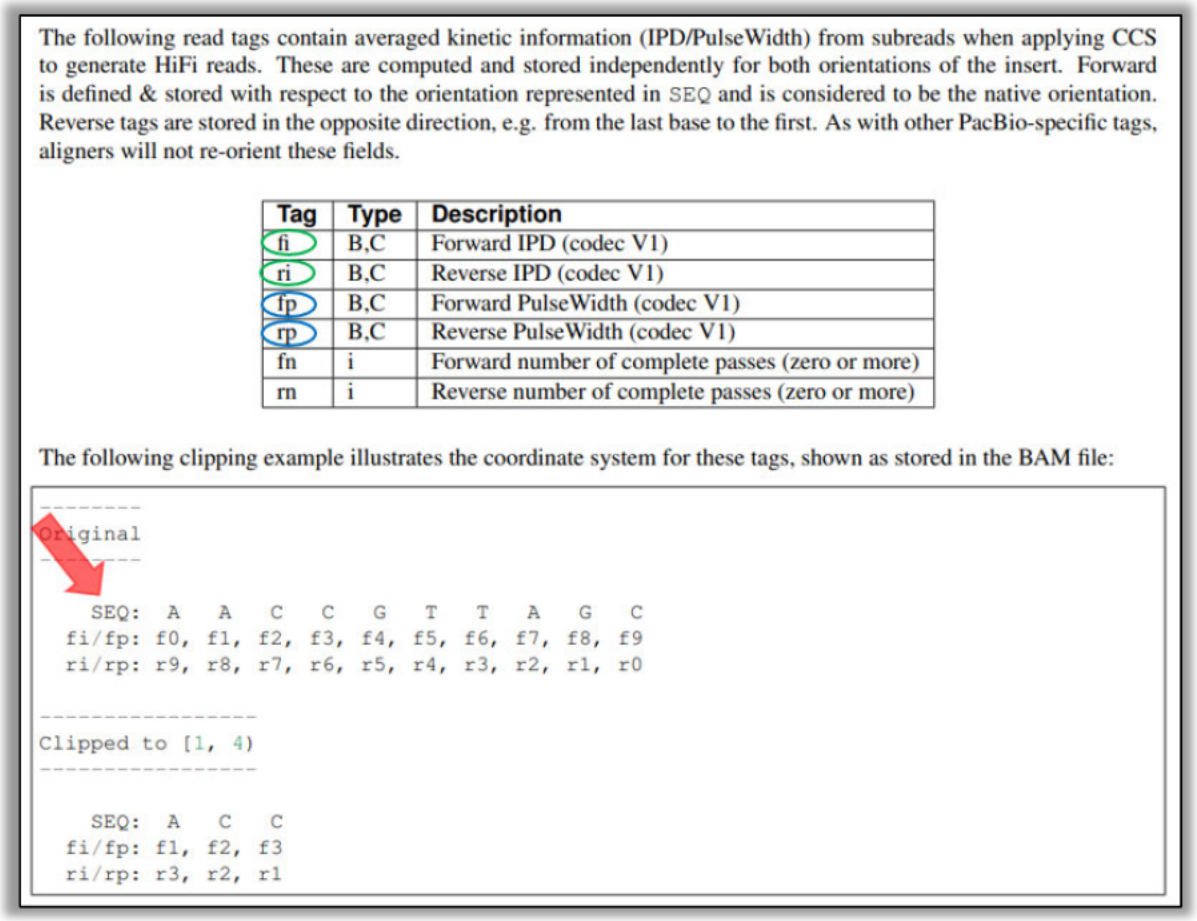


Fig. 5. Screenshot from PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 07:08, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s

51. Thus, the HiFi reads and feature vectors created from the BAM file (i.e., "the input data structure") comprise "a window of the nucleotides sequenced in the sample nucleic acid molecule" as required by claim 1.

52. To detect 5mC methylation, the CCS process analyzes kinetic data as discussed above in paragraphs 43–45, resulting in the creation of a BAM file and feature vectors that include

both conventional and PacBio-specific information. *See* Exhibit G at 5 ("In this document we describe the way we make use of the extensibility mechanisms of the BAM specification to encode PacBio-specific information, as well as conventions we adhere to."). For example, as shown in annotated Fig. 6 below, a BAM file including read tags and feature vectors for HiFi per-read-base kinetic information includes the following properties for each nucleotide within the window of nucleotides for each read or feature vector: (i) the "identity of the nucleotide" is indicated in the `SEQ` field (*see, e.g.*, red arrow), (ii) the "width of the pulse corresponding to the nucleotide" is indicated in the `fp` and `rp` tags (*see, e.g.*, blue circles), and (iii) the "interpulse duration" (or "IPD") is indicated in the `fi` and `ri` tags (*see, e.g.*, green circles).



The following read tags contain averaged kinetic information (IPD/PulseWidth) from subreads when applying CCS to generate HiFi reads. These are computed and stored independently for both orientations of the insert. Forward is defined & stored with respect to the orientation represented in SEQ and is considered to be the native orientation. Reverse tags are stored in the opposite direction, e.g. from the last base to the first. As with other PacBio-specific tags, aligners will not re-orient these fields.

Tag	Type	Description
fi	B,C	Forward IPD (codec V1)
ri	B,C	Reverse IPD (codec V1)
fp	B,C	Forward PulseWidth (codec V1)
rp	B,C	Reverse PulseWidth (codec V1)
fn	i	Forward number of complete passes (zero or more)
rn	i	Reverse number of complete passes (zero or more)

The following clipping example illustrates the coordinate system for these tags, shown as stored in the BAM file:

```
--------
Original
--------

   SEQ:  A   A   C   C   G   T   T   A   G   C
fi/fp: f0, f1, f2, f3, f4, f5, f6, f7, f8, f9
ri/rp: r9, r8, r7, r6, r5, r4, r3, r2, r1, r0

-----------------
Clipped to [1, 4)
-----------------

   SEQ:  A   C   C
fi/fp: f1, f2, f3
ri/rp: r3, r2, r1
```

Fig. 6. Screenshot from BAM format specification for PacBio, available at https://pacbiofileformats.readthedocs.io/_/downloads/en/11.0/pdf/ (annotated) (Exhibit G at 15)

53. Further, each feature vector also includes "a position of the nucleotide with respect to a target position within the window" by aligning the start and end positions of a sequence read (qStart, qEnd, blue circles) to that of a target sequence (tStart, tEnd, green circles) as shown in annotated Fig. 7 below.

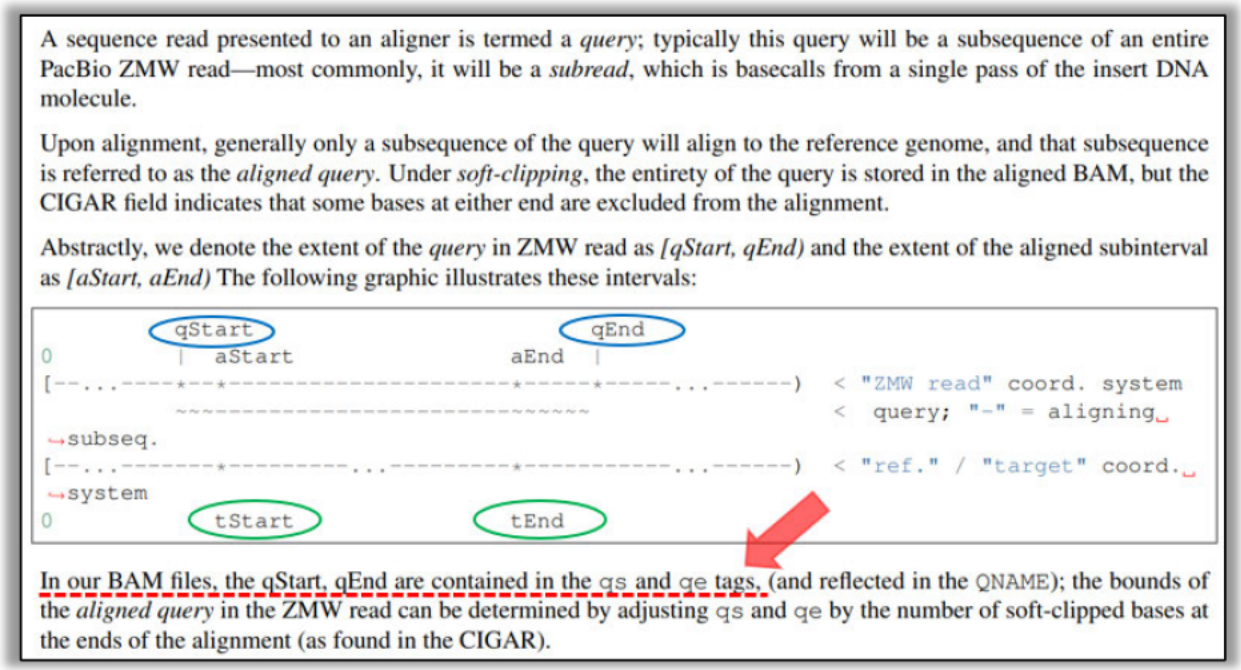


Fig. 7. Screenshot from BAM format specification for PacBio, available at https://pacbiofileformats.readthedocs.io/_/downloads/en/11.0/pdf/ (annotated) (Exhibit G at 71)

54. Consequently, the PacBio Products practice step (b) of claim 1.

55. *Step (c)*. Upon information and belief, the PacBio Products practice step (c) of claim 1 of the '794 Patent, which recites:

> (c) inputting the input data structure into a model, the model trained by:
> receiving a first plurality of first data structures, each first data structure of the first plurality of data structures corresponding to a respective window of nucleotides sequenced in a respective nucleic acid molecule of a plurality of first nucleic acid molecules, wherein each of the first nucleic acid molecules is sequenced by measuring pulses in the optical signal corresponding to the nucleotides, wherein the modification has a known first state in a nucleotide at a target position in each window of each first nucleic acid

> molecule, each first data structure comprising values for the same properties as the input data structure,
> storing a plurality of first training samples, each including one of the first plurality of first data structures and a first label indicating the first state of the nucleotide at the target position, and
> optimizing, using the plurality of first training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels of the first labels when the first plurality of first data structures is input to the model, wherein an output of the model specifies whether the nucleotide at the target position in the respective window has the modification,

56. Upon information and belief, the PacBio Products practice each limitation of step (c) at least by the "primrose" tool first provided in its SMRT® Link Software (v11.0). According to PacBio's SMRT® Tools reference guide (v11.0), the input for the primrose tool is a plurality of feature vectors created or built from a BAM file (i.e., the input data structure described above in paragraphs 48–53). Upon inputting the feature vectors, the primrose tool "uses a convolution neural network (CNN) to predict the methylation state (5mC) of each CpG in a HiFi read." Exhibit J at 103; *see also* Exhibit L at 114. As such, the input data structure is inputted into a model, namely, a CNN model.

57. As discussed in paragraphs 48–53, the BAM file builds a plurality of feature vectors that include "[k]inetic data, pulse width and inter-pulse distance, over a 16 base pair window for both the forward and reverse strand, [which] is used as input to the CNN." Exhibit J at 103; *see also* Exhibit L at 114. During his presentation, Dr. Wenger explains the CNN input further:

> [A]t PacBio we've developed a neural network model to call CpG methylation and HiFi reads. The basic idea is, for every . . . consecutive CG in a read, to produce a feature vector that consists of the kinetics, the pulse widths, and inter-pulse durations in a 16 base window around that CpG site on both strands of the sequence. That's encoded in a vector and then fed into a convolutional neural network model implemented in tensor flow, and that outputs a probability of methylation for that CpG site under the assumption that the methylation is symmetric on the two strands of

> DNA. PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 07:15–08:04, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s.

58. Thus, upon inputting the input data structure (i.e., the feature vector(s)), the model is trained by first "receiving a first plurality of first data structures, each first data structure of the first plurality of data structures corresponding to a respective window of nucleotides sequenced in a respective nucleic acid molecule of a plurality of first nucleic acid molecules, wherein each of the first nucleic acid molecules is sequenced by measuring pulses in the optical signal corresponding to the nucleotides, wherein the modification has a known first state in a nucleotide at a target position in each window of each first nucleic acid molecule, each first data structure comprising values for the same properties as the input data structure[.]"

59. PacBio's SMRT® Tools reference guide (v11.0) further explains that "[t]he CNN is trained using invitro [*sic*] modified controls of methylated and unmethylated human DNA" which correspond to the claimed "plurality of first training samples." Exhibit J at 103; *see also* Exhibit L at 114. According to the reference guide, "[t]he unmethylated control comprises a human shotgun library that has undergone whole genome amplification (WGA), a process that includes PCR, and therefore removes all modifications" while "[t]he methylated control is generated by subjecting a human WGS library to invitro [*sic*] methylation using a CpG methyltransferase enzyme (`M. SssI`)." *Id*. Dr. Wenger further explained these training samples during his presentation:

> [T]he other key attribute to producing a neural network model is training data and for that you want examples of methylation where you have sequences that are either fully methylated or fully unmethylated, and in order to train this model we produced data in a manner where we start with native human DNA for the sample HG002, and in this native DNA of course some of the sites are methylated and some are not. We then performed whole genome amplification using a PCR based approach, which produces then fully unmodified DNA. This library or a library was

> prepared from that DNA is sequenced, and that serves as a set, a large set, genome-wide true negatives of DNA strands that are not modified. Then to produce fully modified sites, we treated this whole genome amplified HG002 DNA with a CpG methyltransferase enzyme, which efficiently adds methylation to any CpG site in the DNA, and this then produces fully modified sequences, which again were sequenced to produce a large set of true positives genome-wide. These two sets were passed into the neural network to train the model.
>
> ***
>
> Moving beyond this . . . synthetically constructed set with either purely modified or purely unmodified sets, we can also apply this model trained in that context to real biological samples, and this has been done in work by the SEQC2 consortium's EpiQC working group, in which they applied PacBio HiFi sequencing and other technologies for detecting methylation to three different reference human materials from genome in a bottle. PacBio, *Methylation Detection with PacBio HiFi Sequencing - ASHG 2021*. YOUTUBE (Oct. 27, 2021) at 08:11–11:02, https://www.youtube.com/watch?v=rT231BCp6Ro&t=20s.

60. Upon information and belief and in view of the Reference Guide and Dr. Wenger's explanation, the CNN model used in PacBio's primrose tool is trained by "storing a plurality of first training samples, each including one of the first plurality of first data structures and a first label indicating the first state of the nucleotide at the target position, and optimizing, using the plurality of first training samples, parameters of the model based on outputs of the model matching or not matching corresponding labels of the first labels when the first plurality of first data structures is input to the model, wherein an output of the model specifies whether the nucleotide at the target position in the respective window has the modification[.]"

61. *Step (d)*. Upon information and belief, the PacBio Products practice step (d) of claim 1 of the '794 Patent, which recites:

> (d) determining, using the model, whether the modification is present in a nucleotide at the target position within the window in the input data structure.

62. As shown in annotated Fig. 8 below, PacBio's website explains that "[a] convolutional neural network model processes polymerase kinetics to determine the methylation

status of each CpG site in a HiFi read."

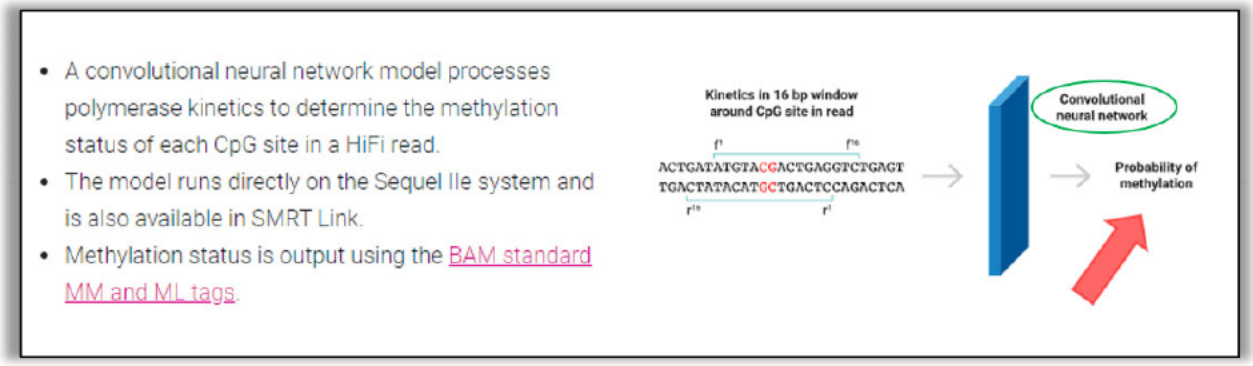


Fig. 8. Screenshot from https://www.pacb.com/products-and-services/applications/epigenetics/ (annotated) (Exhibit F)

63. The SMRT® Tools Reference Guide reference guide (v11.0) further indicates that "[t]he output of the CNN is a probability scale measure of whether the CpG is symmetrically 5mC-modified." Exhibit J at 103; *see also* Exhibit L at 114. In other words, whether the modification (5mC methylation) "is present in a nucleotide at the target position within the window in the input data structure" is determined using the CNN model. That determination is included in "the `Mm` and `Ml` tags as defined in the SAM Format Optional Fields Specification." *Id.*

64. The "method for detecting a modification of a molecule in a nucleic acid molecule" as recited in claim 1 of the '794 Patent, including the individual claim limitations and the invention taken as a whole, was not well understood, routine, or conventional in the art when the '794 Patent was filed. The '794 Patent's claimed methods further improve the length and accuracy of "detection of a modification" over prior art methods. *See, e.g.*, Exhibit A at 20:26–41.

65. PacBio has confirmed that the methods claimed in the '794 Patent, including the individual claim limitations, and the invention taken as a whole, were not well understood, routine, or conventional in the art and further improve the length and accuracy of "detection of a

modification" over prior art methods when explaining the benefits of its infringing products and methods that infringe the claimed invention. *See, e.g.*, Exhibit F at 1 ("Unlike methods that require chemical conversion of DNA, HiFi sequencing detects modifications in Native DNA through impacts on the kinetics of base incorporation" and "High accuracy of sequence and methylation."); Exhibit K at 1 ("Recent improvements in read length and data analysis have extended the ability to include the 5mC methylation that is present at CpG sites in human genomes. Using a deep learning model that integrates sequencing kinetics and base context, the accuracy of 5mC detection in humans for individual HiFi reads is around 80%. Combining multiple reads, the concordance to EMseq and bisulfite sequencing reaches around 90%. The single-molecule resolution of methylation, together with phasing from accurate long reads, allows the detection of allele-specific methylation patterns such as parental imprinting. This ability to detect bases and modifications allows HiFi sequencing to provide the most complete genome and epigenomes with a single technology and library preparation.").

66. PacBio has actively induced, and continues to actively induce, end users of the PacBio Products to directly infringe the '794 Patent. PacBio provides instructions to its customers on how to use its Products to detect 5mC modifications in a manner that infringes the '794 Patent. For example, PacBio's website provides a User Guide for its SMRT® Link software (v11.0, v11.1), a Reference Guide for its SMRT® Tools (v11.0, v11.1) including the CCS analysis and primrose tool, and many resources for customers to learn about the information provided in its BAM files and what that information means. *See, e.g.*, Exhibits D, G, I, J, L, M.

67. Upon information and belief, PacBio had knowledge of the '794 Patent, and of the Plaintiffs' rights therein, on or about August 17, 2021, the '794 Patent's issue date. [redacted]

[redacted]

[redacted]

68. As a direct and proximate result of the foregoing acts of PacBio, the Plaintiffs have suffered, and are entitled to, monetary damages in an amount not yet determined. The Plaintiffs are also entitled to their costs of suit and interest.

69. PacBio's continuing infringement has inflicted and, unless enjoined by this court, will continue to inflict great and irreparable harm upon the Plaintiffs, such as deprivation of the Plaintiffs' rights to exclude others. The Plaintiffs have no adequate remedy at law. The Plaintiffs are entitled to injunctive relief enjoining PacBio from engaging in further acts of infringement.

70. [redacted]. As such, PacBio's continuing acts of infringement are in conscious and willful disregard for the Plaintiffs' rights and the resulting damages to the Plaintiffs is such as to warrant increased damages under 35 U.S.C. § 284 to provide just compensation, and to warrant attorney's fees and costs incurred in prosecuting this action under 35 U.S.C. § 285.

71. PacBio's continuing infringement has inflicted and, unless restrained by this court, will continue to inflict great and irreparable harm upon the Plaintiffs. The Plaintiffs have no adequate remedy at law. The Plaintiffs are entitled to an injunction enjoining PacBio from engaging in further acts of infringement.

JURY DEMAND

72. Plaintiffs demand a jury trial on all issues so triable.

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs respectfully request the following relief:

(a) a judgment that PacBio has infringed and continues to infringe the '794 Patent under at least the provisions of 35 U.S.C. § 271(a) and/or (b);

(b) an order enjoining PacBio and all affiliates, subsidiaries, officers, employees, agents, representatives, licensees, successors, assigns, and all those acting in concert with, or for, or on behalf of PacBio, from infringing the '794 Patent;

(c) a judgment that PacBio's infringement of the '794 Patent is willful;

(d) an award of Plaintiffs' damages or other monetary relief to adequately compensate Plaintiffs for PacBio's infringement of the '794 Patent, and such damages be trebled under 35 U.S.C. § 284 and awarded to Plaintiffs, with pre-judgment and post-judgment interest as allowed by law;

(e) a judgment that this is an exceptional case under 35 U.S.C. § 285 and awarding Plaintiffs their attorneys' fees, expert witness fees, costs, and all expenses incurred in this action, with interest;

(f) an award of all of Plaintiffs' actual and compensatory damages; and

(g) an award of any further and additional relief to Plaintiffs as this Court deems just and proper.

YOUNG CONAWAY STARGATT & TAYLOR, LLP

/s/ Melanie K. Sharp

Melanie K. Sharp (No. 2501)
James L. Higgins (No. 5021)
Taylor E. Hallowell (No. 6815)
1000 North King Street
Wilmington, DE 19801
(302) 571-6600
msharp@ycst.com
jhiggins@ycst.com
thallowell@ycst.com

PERKINS COIE LLP
Michael J. Wise*
Joseph P. Hamilton*
Lara J. Dueppen*
Courtney M. Prochnow*
1888 Century Park East, Suite 1700
Los Angeles, CA 90067-1721
(310) 788-9900

Nathan Kelley*
700 13th Street, NW, Suite 800
Washington, DC 20005-3960
(202) 654-6200

Kyle R. Canavera*
11452 El Camino Real, Suite 300
San Diego, CA 92130-2080
(858) 720-5700

W. Matthew Pierce*
1900 Sixteenth Street, Suite 1400
Denver, CO 80202-5255
(303) 291-2300

**Pro hac vice* admission pending

Attorneys for Take2 Technologies Limited and The Chinese University of Hong Kong

Dated: December 14, 2022

29959915.1